DOCUMENT RESUME

ED 188 460 FL 011 354

AUTHOR Goldsmith, Ross P.; Babcock, Marilyn
TITLE Analysis of the Final Evaluation Reports Submitted to the Bilingual Unit of the Colorado Department of Education by Bilingual Education Programs Funded under the Colorado Bilingual/Bicultural Education Act during the 1977-1978 School Year.
INSTITUTION Dissemination and Assessment Center for Bilingual Education, Austin, Tex.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE [79]
NOTE 79p.; Executive summary, published separately, is attached.
AVAILABLE FROM Dissemination and Assessment Center for Bilingual Education, Educational Service Center, Region XII, 7703 North Lamar, Austin, TX 78752 ($5.00; executive summary $2.00)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Attendance Patterns; *Bilingual Education; Dropout Rate; Educational Objectives; Elementary Education; *English (Second Language); Norm Referenced Tests; Parent Participation; *Program Evaluation; Self Concept; State Programs; Student Attitudes; Student Evaluation; *Student Improvement

ABSTRACT

Results of an evaluation of the Colorado Bilingual Education Program are presented. Data contained in the Colorado Bilingual Education Unit's evaluation reports were analyzed to determine the extent to which the programs have been successful in 1977-78 in meeting the objectives of the Colorado Bilingual Bicultural Education Act. Assessment areas concern the following objectives: improvement in performance in comprehension, reading, writing, and speaking the English language; improvement in school attendance and reduction of the dropout rate; development of a positive self concept and attitudes of the student; and increased parental involvement in the school program. The general study design was to compare how well students did in the program with how well they would have been expected to do without it. Using a norm referenced evaluation design, pre-test scores of participating students were used to generate an expected post-test score against which to compare the observed post-test score of participants. Mean pre-test/post-test scores were used for students within each grade level (K-3) and student classification (linguistically different and nonlinguistically different). Findings indicate that the programs have been generally effective in improving the English reading skills of participating students in the 24 programs reporting some useable pre-test/post-test data. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED188460

ANALYSIS OF THE FINAL EVALUATION REPORTS SUBMITTED TO THE BILINGUAL UNIT OF THE COLORADO DEPARTMENT OF EDUCATION BY BILINGUAL EDUCATION PROGRAMS FUNDED UNDER THE COLORADO BILINGUAL/BICULTURAL EDUCATION ACT DURING THE 1977-1978 SCHOOL YEAR.

by

Ross P. Goldsmith
Assessment & Evaluation Coordinator

and

Marilyn Babcock
Evaluator

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

DACBE

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Assessment Staff:
Ross P. Goldsmith
Evangelina Mangino
Sheila Reifle
Marilyn Babcock

FL011354

Dissemination and Assessment Center for
Bilingual Education
7703 North Lamar
Austin, Texas 78752

TABLE OF CONTENTS

Introduction i

I. Improved performance in Comprehension, Reading, Writing, and Speaking the English Language 4

Figure I-1 - Area under the normal curve divided into NCE's, percentiles, stanines, and standard deviation units. 9

Figure I-2 Distribution of achievement scores for Austin ISD by ethnic group 11

Table I-1 Number of projects reporting gain, loss or no change in English Reading achievement 13

Table I-2 Average pre test scores of reporting projects 20

Table I-2 Average post test scores of reporting projects 23

Table I-4 Project by project pre-post test change. 30-31

II. Improved school attendance and a reduced drop-out rate. 32

Table II-1 Statewide average ratios: Actual days attended to possible days. 34

Figure I-1 Attendance ratios by grade and student classification 35

Table II-2 Attendance ratios by grade, project and student classification 37-38

III. Development of a positive self concept and attitude. 42

IV. Greater parental involvement in the school program. 46

Table IV-1 PAL memberships 47

Table IV-2 PAL activities 49

Table IV-3 Infrequent PAL Activities 50

Table IV-4 Frequent PAL Activities 50

Table IV-5 Other PAL activities 51

Table IV-6 Parent Participation 54

V. Conclusions and Recommendations 53

Appendix A - School Districts submitting evaluation reports. 64

References 67

INTRODUCTION

At the request of the Bilingual Education Unit of the Colorado Department of Education, the Assessment component of the Dissemination and Assessment Center for Bilingual Education at Austin, Texas (DACBE) agreed to review and synthesize the evaluation reports submitted to the Bilingual Unit by the bilingual programs operating under the Colorado Bilingual and Bicultural Education Act during the 1977-78 school year. We (DACBE) agreed to act as external evaluator as a service to the Colorado Department of Education.

The intent of this report is to present the data provided in the evaluation reports submitted to the Department of Education by the bilingual programs and to determine the extent to which the programs have been successful in meeting the objectives as specified in The Act. We have not concerned ourselves with project costs or with any benefits which project participants may have accrued other than those corresponding to the four objectives of The Act. Neither have we addressed such "process" variables as how well the individual projects met their own objectives, how well the needs of the children were assessed, or how closely the projects appeared to have followed prescribed instructional strategies. The entire focus of our effort has been on obtaining as clear and unambiguous an answer as possible to the impact question, "How much more did pupils learn (develop positive self concepts, attend school) by participating in the project than they would have without it?"

Our efforts to determine how much better students did in the bilingual projects than they would have done without them relied on the data contained in the evaluation reports submitted to the Bilingual Unit. This data was often inadequate for our purposes. Although we made the general assumption

that the data presented was factually based, we encountered several instances where apparent errors rendered the data useless. We have accordingly included suggestions and recommendations for improving subsequent evaluations.

The preliminary step in our review of the evaluations involved a thorough reading of each of the original 28 evaluation reports by both of the present authors. After reading each report, the data pertinent to each of the four objectives of the program was summarized on worksheets. The summary and synthesis of the data was spot-checked to determine that no computational errors were introduced in the process. After the data was summarized it was reviewed by both authors prior to preparing a discussion and interpretation of the data. During this process we received an additional eight evaluation reports from the Department of Education. In the event we received an evaluation report after completing our analysis of the data pertaining to one of the objectives of The Act, the data included in that late report was not included in the final analysis. The only exception to this rule is with regard to data pertaining to student achievement in English reading. Any data that arrived late was included in the analysis for that objective of The Act. As a result, the data discussed with regard to each objective of The Act is based on slightly different numbers of project reports. The number of project reports included in the data pertaining to any one objective is clearly stated in the discussion of that objective.

The body of this report is divided into sections that correspond to each of the four objectives of The Act. The final section of the report includes a summary of our findings and recommendations for improvement in the evaluation procedures for bilingual programs in Colorado.

We have deliberately avoided identifying specific districts or projects and the data they individually reported. This was acoomplished by assigning a

number to each evaluation report. In this way, data for individual projects could be presented without identifying the district by name. This allows the reader to compare, e.g., attendance ratio of Kindergarten L.D. children in project #25 with the achievement data reflecting whether these children gained, lost or held their own between pre- and post-testing. Where data is presented in this project by project form, the reader will note that no entry appears for some projects (e.g., project #4). The absence of such an entry indicates that that project either did not report the requested data or that the reported data was judged unusable and was therefore excluded from analysis.

Throughout the body of this report, the use of the pronoun "we" refers to the authors of this report unless otherwise specified. The views expressed in this report remain the responsibility of the authors. They do not necessarily reflect the positions or policy of DACBE, its grantor (the U.S. Office of Education) or the grantee, Education Service Center XIII.

IMPROVED PERFORMANCE IN COMPREHENSION, READING, WRITING, AND SPEAKING THE ENGLISH LANGUAGE

The term "improved" as used in this objective of the act, implies a comparison between the performance observed of students in the bilingual program with some estimate of what their performance would have been in the absence of the bilingual program. It would not be meaningful to define "improved" as simply evidence that the post-test scores of participating students are higher than were their pre-test scores. The flaw in such an approach is that it assumes that no learning would be expected to occur between pre- and post-testing if it were not for the particular educational program the child is in. It is a generally accepted truism in educational evaluation that post-test scores will be significantly higher than pre-test scores if there is a sufficient interval between testing. Such an observed gain could be ascribed to maturation, or to practice effect, or to both, and can be independent of the effect of the educational program itself. Such an approach does not answer the question "How much better do students do in the project than they would have done without it?". Rather, it answers the question "Did the students learn anything at all?" Since it is difficult to prevent students from learning something over a period of time, answering this question is not useful.

In order to meaningfully determine whether children in the program learned more than they would have had they not been in the bilingual programs, some estimate of the learning rate they would be expected to evidence in the absence of any special program must be generated. The most appropriate method of doing this is to utilize a control group of students who do not receive the bilingual program. Such a control group must be comprised of students that are, for all intents and purposes, identical to the students that are in the

experimental program. None of the evaluation reports reviewed included such a control group. In the absence of such control groups, the task of estimating the expected rate of growth in the absence of the program becomes somewhat more difficult. A comparison could be made with the rate of increase in the same group of children before they entered the bilingual program. No such data was available. A comparison could be made with the increases observed in previous groups of students that did not participate in the program, but would have been eligible to have done so. No such data was available. The final alternative is to generate some reasonable expectation of what the learning rate would be in the absence of a bilingual educational program. This procedure involves the use of norms accompanying standardized achievement tests. We elected to follow this option.

Making use of a norm-referenced evaluation model (Horst, Tallmadge and Wood, 1975) requires that students have both a pre- and post-test score. The pre-test score is then used to generate the EXPECTED post-test score against which the OBSERVED post-test scores are compared. Simply stated, we would expect a child scoring at the 35th percentile on pre-testing to score at the 35th percentile on post-testing. To accomplish this, that child would have to learn something since all the rest of the children against whom he is being compared have learned something between testings. In order to maintain his relative status with respect to the norm group some learning must have taken place. In the event that an unusual amount of learning had taken place then we would expect the child's post-test score to be in excess of the 35th percentile. If this state of affairs was observed, we might conclude (depending on the extent to which observed post-test score differs from the expected post-test score) that the child had evidenced more learning than would have been expected in the absence of some unusually effective educational intervention.

Since the application of this model requires that pre- and post-test scores be available, we reviewed all evaluation reports and identified those projects for which a pre- and post-test score was reported for Linguistically Different students in at least one grade. Twenty-four projects reported usable pre- and post-test data at at least one grade level for Linguistically Different students. It should be understood that few projects reported usable pre- and post-test data for all grades (K-3). Projects that reported data from tests that were not normed were not included in this analysis. Likewise projects that reported the results of criterion-referenced tests were not treated.

Most projects reported pre- post-test scores in terms of the average raw score of the children in the program. A few reported the scores in terms of the average percentile, and some reported the score in terms of grade equivalents. None of these metrics are usable for our purpose. Therefore, all scores were converted to Normal Curve Equivalents. This process involved going to the norm tables of each test and looking up the percentile rank that corresponded to the reported raw, or grade equivalent score and then converting the percentile rank to a Normal Curve Equivalent (NCE). In this process we discovered that at least one project had apparently committed an error in reporting the name and form of the test used, or had committed a mathematical error in computing the score. When we referred to the norm tables of the test the project claimed to have used, we discovered that the scores they reported were impossible: their average raw score exceeded the total possible. This data was discarded.

It was necessary to convert the reported scores to NCEs so as to generate scores that comprised equal interval scales. That is, the units of the scale must be of the same size over the entire range of the scale values so that any gain would represent exactly the same amount of growth regardless of whether

it occurred at the low end of the scale, the middle of the scale, or the high end of the scale. NCE's are one type of normalized standard scores. Standard scores are simply scores that express the difference between a particular "observed" score and the mean score of the total group (in this case, the norm group of the test). Standard scores are expressed in standard deviation units. In this sense NCE scores simply reflect the mean scores of the individual projects expressed as deviations from the national average. NCE's have a mean of 50 and standard deviation of 21.06. Standard scores have the advantage of being equal interval scales as well as being related to the percentile rank of the distribution of scores. As such it is possible to treat them mathematically as well as make meaningful interpretations of the scores themselves in terms of their relationship to the norm. Neither grade equivalents nor percentiles are equal interval scales, therefore they can not be meaningfully added, subtracted, or multiplied together. Likewise, it is not possible to equate a gain of 5 percentile points, for example, from the 15th to the 20th percentile with a gain of 5 percentile points from the 50th to the 55th percentile. Not only do grade equivalent score have the same disadvantage as percentile scores (they are not equal interval scales) they have a number of other, in some ways more serious, technical weaknesses as well. They should never be used in education evaluation.

Since standard scores such as NCE's directly correspond to a percentile rank within the distribution of scores, the NCE representing the pre-test score can be used as the expected post-test score. It is then possible to simply examine the difference between the NCE obtained at pre-testing and the NCE obtained at post-testing to determine whether a meaningful educational gain has been made. Ideally, one would perform a statistical test (a t-test for correlated means) between the observed post-test NCE's and the expected

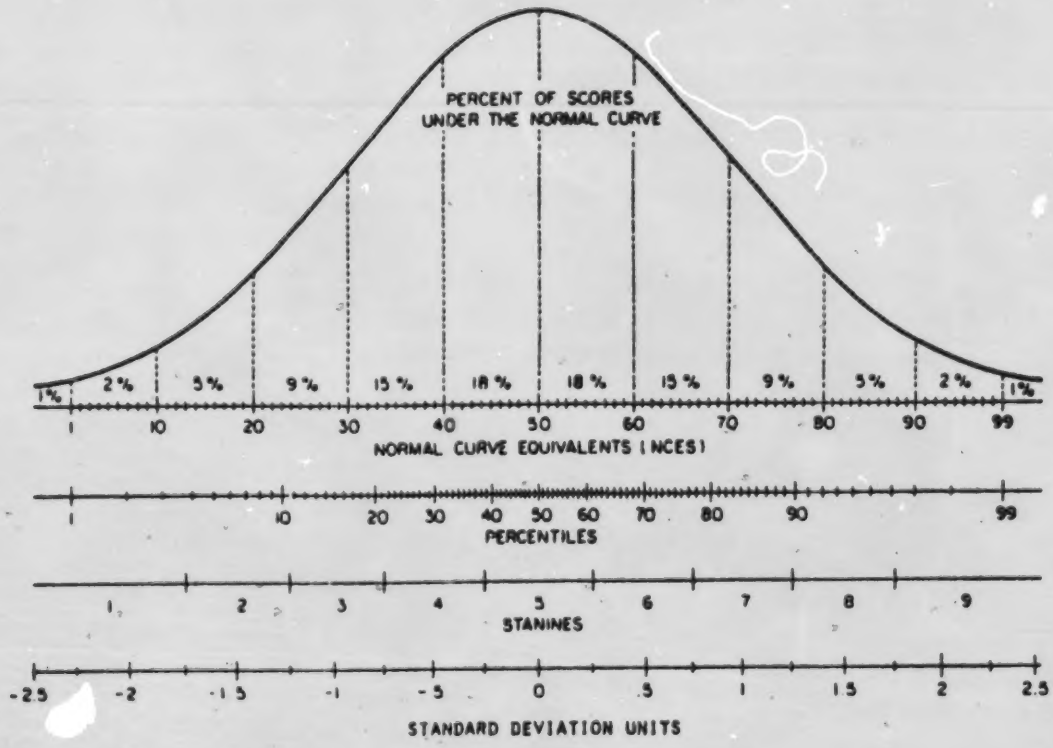
post-test NCEs. Such a procedure requires more data than just the mean pre- and post-test scores. One must know the number of students (N) represented in the pre- and post-test means as well as the standard deviations of the pre- and post-tests. This data was not reported. Therefore, no statistical tests of the significance of the difference between the pre- and post-test NCEs could be accomplished.

It was, therefore, necessary to establish a criterion for judging when a significant gain (or loss) had occured. A gain of one third of the standard deviation with respect to the national norm was chosen as the criterion to be used in making this judgment. Since the standard deviation of NCEs is, by definition, 21.06, a significant gain (or loss) was judged to have occured if the post-test score was seven or more NCEs above or below the pre-test score for any grade level group of students. This criterion was selected on, admittedly, somewhat arbitrary grounds. It has been recommended as a criterion for judging educational (as opposed to statistical) significance of educational programs (Horst, Tallmadge and Wood, 1975). It has been used in a number of federally funded studies of educational program impact. It is also used as a criterion for identifying exemplary programs for packaging and dissemination by the Joint Dissemination Review Panel of the Education Division of the Department of Health, Education, and Welfare.

The reader should appreciate that this criterion is a very stringent requirement of an educational program. Examination of Figure 1 will disclose that an increase of seven NCEs is substantial. For example, if a child has a pre-test score of NCE = 50, and a post-test NCE = 57 this means that the child has moved from the 50th percentile to the 63rd percentile. A gain of this magnitude is extraordinary. In fairness, the reader should also note that a gain of seven NCEs is not necessarily a gain of thirteen percentile points, since percentiles do not constitute an equal interval scale. If a child's

FIGURE I-1

Area under the normal curve divided into NCEs, percentiles, stanines, and standard deviation units.



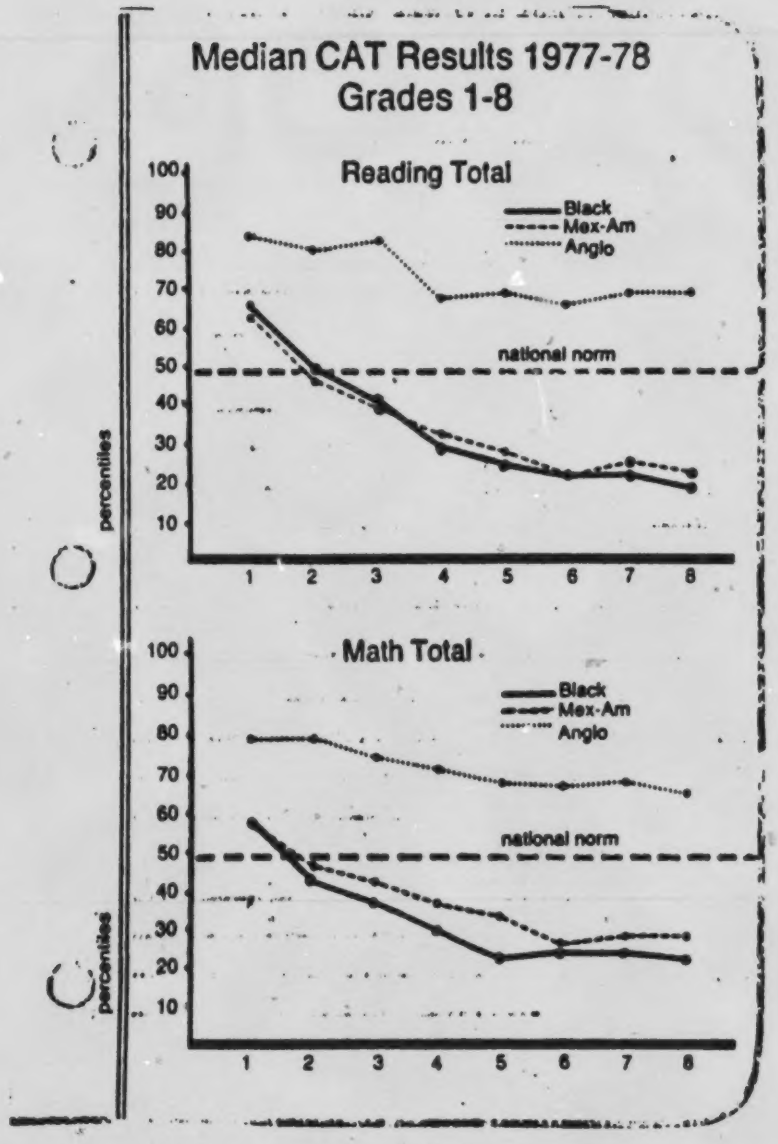
pre-test NCE was 20 and the post-test NCE was 27, the observed difference would represent a gain of six percentile points. Although the improvement in percentile standing is different in each of these examples, the amount of gain is the same.

It is clear, then, that if the Linguistically Different students in the Kindergarten of a bilingual program evidenced a gain of seven or more NCEs that the students did better in the project than we estimate that they would have done without the project. We would judge that such a project had been successful in meeting the objective of The Act for the Linguistically Different students in Kindergarten. What may be somewhat less clear is that if the Linguistically Different students in such a program evidenced no gain, but did not evidence a loss, that the program might be judged to be successful in preventing a deterioration of the expected learning rate of the students. There is a large body of literature that suggests that Mexican-American students (and most Linguistically Different students in Colorado Bilingual Programs are Mexican-Americans) do not maintain the same post-test status with regard to pre-test status through the public school process. Rather, this group of children seems to lose standing with respect to national norm groups. This is illustrated in Figure I-2 which displays the actual achievement scores in percentiles, for the Austin public schools during the 1977-78 school year. This graph is broken down into ethnic groups. Examination of this graph will reveal that the Mexican-American (and Black) students in this district are at the national mean in Kindergarten and fall further and further below the national mean as they progress through the grades. Similar trends have been documented in many other communities (for example Coleman, 1966).

Technically, the hypothesis we are testing with the gain of seven NCE procedure noted above is that, in the absence of a bilingual educational

FIGURE I-2

Distribution of Achievement Test Scores
for Austin ISD by Ethnic Group



program, the students will stay at the same level on post-test that they evidenced on pre-test. The discussion above highlights the fact that this hypothesis may not actually represent the true educational expectation for these children. There is at least some reason to believe that a more reasonable expectation of Linguistically Different students is that they will fall progressively further and further behind, rather than holding their own. In light of this analysis, a program that prevents such students from falling further and further behind as they progress through the grades would have to be judged successful, at least in part. Therefore, we would argue that programs in which Linguistically Different students evidenced neither a significant gain nor a significant loss are, at least in part, successful. Educators would do well to join with physicians in vowing that they will "in the first instance, do no harm."

Table I-1 displays the number of projects for each grade level in which the Linguistically Different children and the Non-Linguistically Different children showed either a gain or a loss or no change in their pre- post-test status. In Kindergarten, 13 projects reported usable pre- post-test data for the Linguistically Different students. One project data evidenced that the Linguistically Different students had a significant loss from pre- to post-testing. In five projects, the Linguistically Different Kindergarteners showed no change from pre- to post-tests. Seven projects reported that their Linguistically Different students evidenced a significant gain in Kindergarten. Twelve projects reported usable pre- post-test data for Non-Linguistically Different students in the bilingual program at the Kindergarten level. One such project reported that the Non-Linguistically Different students lost ground, three projects reported that the Linguistically Different students evidenced no change and eight projects reported that the Non-Linguistically Different students evidenced a significant gain.

TABLE I - 1

Number of Projects Reporting Significant Gains, Losses or No Change in English Reading Achievement by Grade Level and Student Classification.

Student Classification	Kindergarten		
	Lost 7 or more NCEs	No Change (Pre & Post tests within 6 NCEs)	Gained 7 or more NCEs
L.D. n=13	1	5	7
NLD n=12	1	3	8
Student Classification	**First Grade**		
	Lost 7 or more NCEs	No Change	Gained 7 or more NCEs
L.D. n=11	3	3	5
NLD n=11	1	5	5
Student Classification	**Second Grade**		
	Lost 7 or more NCEs	No Change	Gained 7 or more NCEs
L.D. n=18	1	13	4
NLD n=16	1	10	5
Student Classification	**Third Grade**		
	Lost 7 or more NCEs	No Change	Gained 7 or more NCEs
L.D. n=14	1	9	4
NLD n=12	0	9	3

n = number of projects reporting usable pre-post test data for specified grade level group of students

L.D. = Linguistically Different students

NLD = Non-Linguistically Different students

Eleven projects provided usable pre- post-test data with regard to the Linguistically Different students in the first grade bilingual programs. Three of these projects reported that their first grade Linguistically Different students evidenced substantial loss. Three projects reported that the Linguistically Different students experienced no significant change and five projects reported that the Linguistically Different students experienced a substantial gain. Eleven projects reported usable pre- post-test data for first grade Non-Linguistically Different students in the bilingual program. One of these projects reported that their Non-Linguistically Different students experienced a loss, five of the projects reported that their Non-Linguistically Different students experienced no significant change and five reported that these students achieved a significant gain.

Eighteen projects reported usable pre- post-test data for their second grade Linguistically Different students participating in the bilingual program. One such project reported that these children experienced a significant loss. Thirteen of the projects reported that their Linguistically Different students experienced no significant change between pre- and post-testing and four projects reported a significant gain for the second grade Linguistically Different students. Sixteen projects reported usable data for their Non-Linguistically Different students in the bilingual program at second grade. One of these reported a significant loss, ten reported there was no change, and five reported that the Non-Linguistically Different students experienced a significant improvement.

Fourteen projects reported usable data for the Linguistically Different students in their third grade programs. One of these reported that their Linguistically Different students experienced a loss, nine reported that the Linguistically Different students evidenced no change. Four reported that

they experienced a significant gain. Twelve projects reported usable data for Non-Linguistically Different students in the third grade projects. None of these projects reported that the Non-Linguistically Different students experienced a loss, nine reported that they experienced no change, and in three projects Non-Linguistically Different students evidenced a significant gain.

At this point it is appropriate to discuss a few of the limitations of the conclusions drawn above. First, the pre- post-testing that was accomplished at the Kindergarten level made use (by necessity) of tests that purport to measure reading readiness rather than reading achievement. This is, of course, appropriate since Kindergarteners are not expected to have learned how to read and, therefore, it would not be appropriate to test their achievement in that area. However, reading readiness tests are of relatively limited value. Children typically obtain higher scores on such measure than their subsequent academic performance in the area of reading achievement would appear to justify. In other words, a child might appear to be eminently ready to read, and yet experience only indifferent success in subsequent reading achievement. As a result, the apparent gains and losses noted for Kindergarten may well bear little relationship to subsequent performance in school.

A second concern we have about the pre-post-test data, at all grade levels, is that the number of children represented in the pre-test and in the post-test was rarely reported. For a pre-post-test evaluation design to be meaningful, it is essential that the same children that were pre-tested also be those children that are post-tested. Where the make-up of the group changes from pre- to post-testing, it is impossible to make any judgments about the effect of the program on its participants, since one is making a

comparison between the pre-tests of one group of children and the post-tests of another. Since this information was usually not included in the evaluation reports, we were forced to assume that the children that took the pre-test were also the children that took the post-test. Skeptical evaluators (among whose numbers we count ourselves) are very unwilling to make such an unqualified assumption. One of the few projects that reported the number of students in pre- and post-testing noted that 60 Linguistically Different students were included in the pre-test, but only 12 in the post-test. This level of attrition renders any pre-post-test judgments meaningless. Accordingly, the data from that project is not included in this analysis. A general rule is that if as few as 10 percent of program participants are omitted from the pre- or post-test, the resulting data may be seriously compromised and any conclusions based on such data may be spurious. Where the reports included data on the number of children present for pre- and post-testing, we were able to judge whether an unacceptable level of attrition was evident. When it was, the data was excluded from analysis. Where the N was unreported, we had no way of determining whether undesireable levels of attrition had occured. In the absence of any evidence to the contrary, we had to assume that such attrition had not occured, and consequently we included these projects in the analysis.

This observation leads us to strongly recommend that subsequent evaluation reports from Colorado Bilingual Projects be required to specify the number (N) of students represented in each of the testings reported. We also strongly suggest that the standard deviation of the distribution of each testing be reported, along with the mean as is currently done. We also strongly recommend that districts report their scores in a uniform metric. We prefer that that metric be specified as NCEs. It takes a great deal of time and effort to make sense out of any other sort of score.

The third concern we have is that in several instances projects used one test for the pre-test and another test for the post-test. It is virtually never possible to make use of scores from different tests in judging pre-post-test differences. This was most commonly encountered in first grade where the pre-test was a reading readiness test (such as the TOBE) and the post-test a reading achievement test (such as the CAT). Pre-post-test data of this sort was excluded from this analysis.

The fourth concern is that one project made use of criterion-referenced tests as the exclusive measure of the impact of the project on the student's academic achievement. Criterion-referenced tests are quite useful for teachers in the day-to-day instructional process. In the summative evaluation of an educational program criterion-referenced testing is useful only if: 1) common objectives are set for all schools, 2) the scoring system is objective, and 3) the standards set are tied to agreed-upon typical performances and well-defined conditions instead of arbitrary levels of performance and conditions. These prerequisites are almost never met in program evaluation with criterion-referenced tests. These criteria were not met by any of the Colorado projects that employed criterion-referenced testing. As a result it was impossible to form any judgment based on criterion-referenced data as to whether the students in those projects did better than they would have done without those projects. These projects reported that virtually all of their students met virtually all of their objectives as measured by their respective criterion-referenced tests.

We would urge districts to make use of standardized, norm-referenced achievement tests in the summative evaluation of their educational programs. Criterion-referenced tests have often been touted as the alternative to many of the undesireable manifestations of standardized testing. Although

standardized tests are not perfect, they are vastly superior to currently available alternatives for summative evaluation purposes.

The astute reader may well wonder how likely it is that the gains noted above might have occurred by chance. We are unable to make any firm statement about the probability of the observed pre-post-test differences occurring simply by chance, because the data that would be necessary to make such a determination (the N and standard deviation for each mean and the correlation between the pre- and post-test scores for each group) were not reported. Without this information it is impossible to perform the statistical tests that would determine the likelihood that the observed differences would have occurred by chance alone. It should be noted, however, that pre-post-test differences considerably smaller than seven NCE's (the standard we have used to judge educational significance) may be found to be statistically significant. This illustrates a general artifact of statistics -- if the number of subjects is large, then the differences between their scores may be quite small and yet be statistically significant. When an educationally significant gain (or loss) is observed, where our definition of a significant gain or loss is used, then that difference will usually prove to be statistically significant. The reverse is not necessarily true; differences are frequently statistically significant but not educationally significant.

Since not all Colorado projects reported usable pre-post-data the reader may well ask whether the results of the projects that did report usable data are generalizable to the projects that did not. Unfortunately, such a determination can not be made.

Many projects reported either pre- or post-test data from one or more grade levels with respect to Linguistically Different or Non-Linguistically Different students in the program. Although these projects did not report

usable pre-post-test data they have some apparently valid data points that can be described. Table I-2 displays the average pre-test data reported for Colorado bilingual projects. This table categorizes the average score of project participants by the percentile ranges in which they fall. We have elected to divide the percentile ranges (on somewhat arbitrary grounds) into four general ranges: the 37th percentile and below, the 38th through 49th percentile, the 50th through 62nd percentile, and the 63rd percentile and above. We selected these percentile cut-off scores since they correspond to ranges within plus or minus seven NCEs of the mean (50th percentile). This system of categorization has the advantage of corresponding to the criteria that we adopted in judging educationally significant gains or losses (plus or minus seven NCEs). The data reported in Table I-2 includes the pre-test scores for those projects where usable pre- and post-test data was reported as well as those projects where only usable pre-test data for any given grade level was reported.

As noted in Table I-2, five projects (of 13 reporting usable data) reported that their Kindergarten pre-test for Linguistically Different students averaged at or below the 37th percentile. Three projects reported that their Linguistically Different students scored on the average between the 38th and 49th percentile on pre-test, four reported that the Linguistically Different students scored between the 50th and 62nd percentile, and one reported that the Linguistically Different students scored at or above the 63rd percentile at pre-testing. Twelve projects reported usable pre-test data for Non-Linguistically Different students in Kindergarten. The average pre-test score for three of these projects was at or below the 37th percentile, one project's students scored on the average, between the 38th and 49th percentile, two projects' Non-Linguistically Different students scored between the 50th and 62nd percentile and six projects

TABLE I-2

Number of Projects Reporting Average PRE TEST Scores Within Specified Percentile Ranges by Grade Level and Student Classification

Student Classification	At or Below the 37th Percentile	Between the 38th and 49th Percentiles	Between the 50th and 62nd Percentiles	At or Above the 63rd Percentiles
Kindergarten				
L.D. n=13	5	3	4	1
MLD n=12	2	1	2	6
First Grade				
L.D. n=13	6	2	3	2
MLD n=13	3	3	3	4
Second Grade				
L.D. n=18	5	5	3	5
MLD n=17	2	2	6	7
Third Grade				
L.D. n=14	4	5	4	1
MLD n=12	1	1	4	6

n = Number of projects reporting usable pre test data for specified grade level and student classification

reported that the average of their Non-Linguistically Different students' pre-test score was at or above the 63rd percentile.

Thirteen projects reported usable pre-test for Linguistically Different students in first grade. Six of these projects reported that the average pre-test score for Linguistically Different students in first grade was at or below the 37th percentile, two projects reported average Linguistically Different pre-test scores between the 38th and the 49th percentile, three reported that their scores were between the 50th and the 62nd percentile, and two reported that their scores were at or above the 63rd percentile. Thirteen projects reported usable pre-test data for Non-Linguistically Different students in first grade. Three of these reported their average score as at or below the 37th percentile, three reported the average between the 38th and 49th percentile, three reported an average between the 50th and 62nd percentile, and four reported that their Non-Linguistically Different students average score was at or above the 63rd percentile.

Eighteen projects reported usable pre-test data for Linguistically Different students in the second grade of the programs. Five of these projects reported the pre-test scores for such children averaged at or below the 37th percentile, five reported the pre-test scores averaging between the 38th and 49th percentile, three between the 50th and 62nd percentile, and five reported Linguistically Different students' average pre-test scores at or above the 63rd percentile. Seventeen projects reported usable pre-test scores for Non-Linguistically Different students in the second grade of the program. Two of these projects reported that these students' average pre-test score was at or below the 37th percentile, two projects reported their Non-Linguistically Different students scored between the 38th and 49th percentile, on the average, six reported that their average Non-Linguistically Different student pre-test

score was between the 50th and 63rd percentile, and seven projects reported that their Non-Linguistically Different student pre-test score average met or exceeded the 63rd percentile.

Fourteen projects reported usable pre-test data for Linguistically Different students in the third grade of the bilingual programs. Four of these projects reported the average pre-test score for Linguistically Different students was at or below the 37th percentile, five reported average scores between the 38th and 49th percentile, four reported scores between the 50th and 62nd percentile, and one project reported that Linguistically Different students pre-tested, on the average, at or above the 63rd percentile. Twelve projects reported usable pre-test data for their Non-Linguistically Different students in third grade of the program. One of these projects reported that these students' average pre-test was at or below the 37th percentile and one project reported that Non-Linguistically Different students pre-tested between the 38th and the 49th percentile, on the average. Four projects reported the Non-Linguistically Different pre-test scores averaging between the 50th and 62nd percentile and six projects reported average pre-test scores for Non-Linguistically Different students at or above the 63rd percentile.

Table I-3 displays the post-test averages for projects in Colorado. This Table makes use of the same classification system as used for pre-test scores.

As noted in Table I-3, eighteen projects provided usable post-test scores for Linguistically Different students in the Kindergarten program. None of these projects reported that the Linguistically Different students post-tested at or below the 37th percentile. Five reported average post-test scores for Linguistically Different students between the 38th and 49th percentile, five reported average post-test scores between the 50th and 62nd percentile and eight reported that their Linguistically Different students average post-test

TABLE I-3

Number of Projects Reporting Average POST TEST
Scores Within Specified Percentile Ranges by
Grade Level and Student Classification

Student Classification	At or Below the 37th Percentile	Between the 38th and 49th Percentiles	Between the 50th and 62nd Percentile	At or above the 63rd Percentile
Kindergarten				
L.D. n=18	0	5	5	8
NLD n=16	0	2	2	12
First Grade				
L.D. n=27	7	6	9	5
NLD n=24	0	2	8	14
Second Grade				
L.D. n=22	5	8	5	4
NLD n=19	0	2	3	14
Third Grade				
L.D. n=18	3	8	5	2
NLD n=17	0	2	3	12

n = Number of projects reporting usable post test data for specified grade level and student classification

score was at or above the 63rd percentile. Sixteen projects reported usable post-test data for Non-Linguistically Different students in the Kindergarten level of the program. None of the projects reported that their Non-Linguistically Different students scored at or below the 37th percentile on post-testing. Two reported their Non-Linguistically Different students scoring between the 38th and 49th percentile at post-testing and two reported that the average of their Non-Linguistically Different students post-test scores fell between the 50th and 62nd percentile. Twelve projects reported that their average post-test score for Non-Linguistically Different students in Kindergarten was at or above the 63rd percentile.

Twenty seven projects reported usable post-test data for Linguistically Different students in first grade. Seven of these projects reported that their Linguistically Different students averaged below the 38th percentile on post-testing. Six reported that Linguistically Different students averaged between the 38th and 49th percentile, nine reported that Linguistically Different students post-tested between the 50th and 62nd percentile, and five reported that their Linguistically Different students' average post-test score was at or above the 63rd percentile. Twenty four projects reported usable post-test data for Non-Linguistically Different students in first grade. None of these projects reported that their Non-Linguistically Different participants averaged at or below the 37th percentile. Two reported that these students' average post-test scores fell between the 38th and 49th percentile, eight reported that the average post-test score was between the 50th and 62nd percentile, and fourteen projects reported that the Non-Linguistically Different students post-test scores averaged at or above the 63rd percentile.

Twenty two projects reported usable post-test data for their second grade Linguistically Different students. Five of these projects reported that their

Linguistically Different students' post-test scores averaged at or below the 37th percentile. Eight reported that the Linguistically Different students' average post-test score fell between the 38th and 49th percentile, five reported that these students' average post-test score fell between the 50th and 62nd percentile and four reported that the Linguistically Different students' post-test scores fell at or above the 63rd percentile. Nineteen projects reported usable post-test data for the Non-Linguistically Different students in their second grades. None of these projects reported that Non-Linguistically Different students' post-test scores were at or below the 37th percentile. Two projects reported average Non-Linguistically Different post-test scores between the 38th and 49th percentile, three between the 50th and 62nd percentile, and fourteen at or above the 63rd percentile.

Eighteen projects reported usable post-test scores for Linguistically Different students in the third grade of their programs. Three of these projects reported average post-test scores for Linguistically Different students at or below the 37th percentile, eight reported average post-test scores for Linguistically Different students between the 38th and 49th percentile. Five projects reported that Linguistically Different students' average post-test score fell between the 50th and 62nd percentile, and two projects reported that Linguistically Different students scored at or above the 63rd percentile, on the average at post-testing. Seventeen projects reported usable post-test data for Non-Linguistically Different students in the bilingual program. None of these projects reported that the average post-test score was at or below the 37th percentile. Two reported their Non-Linguistically Different students' post-test scores averaged between the 38th and 49th percentile, three between the 50th and 62nd percentile and twelve at or above the 63rd percentile.

The reader should note that these pre- and post-tests do not represent all of the bilingual projects operating in the state. A number provided no usable pre- or post-test data for one, several, or all grade levels or student classifications. Unfortunately the largest project in the state failed to separate their test scores into scores attributable to LD students and those attributable to Non-LD students. As a result the data from this project was entirely unusable in this or the previous analysis.

Close examination of the post-test scores as displayed in Table I-3 reveals that the average scores for the Non-LD students in the projects reporting usable data usually fell within respectable limits, at or above the national norm (the 50th percentile). Although it is impossible to determine whether the program is responsible for the achievement status of these children based on post-tests results alone, it would appear that they have not suffered by virtue of their participation in the bilingual program. These post-test scores also suggest that the average performance of Linguistically Different students in many projects is at or above national norms. Once again, it is impossible to determine whether this performance is a result of their participation in the bilingual program since not all projects reporting post-test scores also reported usable pre-test scores.

These pre- and post-test scores should not be interpreted to mean that all of the students in the programs are scoring within the ranges we have specified here. This data only displays what the average score of children in each of those projects has been reported to be. It must be assumed that some of the children scored below the average of their grade level and classification group within each project and that some scored above that group average. Once again, due to the absence of pertinent data, it is impossible to make any statements about how many of the participants included in any project average might be scoring well above or well below that average point.

Finally, it must be stressed that the scores represented here reflect the average score of the students participating in a project. The fact that three fourths of the projects reporting usable Kindergarten post-test data for Non-Linguistically Different students found that the average post-test was at or above the 63rd percentile should not be interpreted to mean that three fourths of the Non-Linguistically Different students in Kindergarten bilingual programs are scoring at or above this level. These figures represent the averages reported per project and do not reflect the number of children that are represented by these figures. It is impossible to determine what proportion of (for instance) the Non-LD students participating in Kindergarten bilingual programs in Colorado are scoring at or above a certain level. The number of children participating in the program varies considerably from one project to another. Therefore the average pre- or post-test score for one project may represent the scores of twenty children, while the average pre- or post-test score in another may represent several hundred children. We are unable to determine conclude how many children this data represents from the available information.

We would be remiss if we failed to mention one final limitation of the available data and our treatment of it. When a norm-referenced evaluation model is used to judge the impact of an educational program, the test data should be collected at the same time during the school year that the test publisher collected their normative data. Although test publishers typically provide norm tables for Fall, Mid-year, and Spring they have often only collected data for one point during the year (most frequently at the end of February or the first of March). Therefore, their Fall and Spring norm tables represent extrapolations rather than real data. When norm tables based on estimates of what students would have scored at that time in the year are used

in norm-referenced evaluations, they may introduce an undesireable level of measurement error and therefore result in evaluation findings that do not accurately represent the true state of affairs. A number of tests with just one normative data point were used in pre-post-testing in Colorado Bilingual Programs. We feel that it is essential that subsequent evaluations of Colorado Bilingual Programs employ tests, and testing schedules, that correspond to the dates when the norms for the tests employed were actually collected. Following this recommendation will considerably strengthen the validity of the evaluation conclusions.

Among the standardized, norm-referenced, achievement tests used in Colorado evaluations, the California Achievement Test (CAT), the Comprehensive Tests of Basic Skills (CTBS) and the Iowa Test of Basic Skills (ITBS) all provide norm tables based on data collected at only one time during the academic year. For the CAT and CTBS, actual normative data was collected at the end of February and the first part of March. For the ITBS, norm data was collected during the last half of October and the first half of November. If these tests are to be used for pre-post-testing, the testing should be conducted only once a year and within two weeks of the actual dates that the norm data was collected by the publishers. The Metropolitan Achievement Test (MAT), the Gates McGinty, and the Stanford Achievement Test (SAT) all collected their normative data at several times during the year. The MAT and Gates McGinty collected their data during October and April. The SAT collected normative data during October, February and May. These tests can be used for pre-post testing (Fall to Spring). Nevertheless, care should be exercised to conduct testing at the same time that the test publishers collected their norm data.

Summary

It is the position of the authors that conclusive generalizations about an educational program are possible only after amassing consistent evidence from a variety of evaluations over a period of time. No single try-out can provide a sound basis for generalizations, no matter how carefully it is conducted. The evaluation of educational programs is an extremely difficult task. Decision makers must understand that good evaluations are not easy to accomplish. Sound educational evaluation requires adequate resources and, most importantly, competent evaluators. The former are much easier to obtain than the latter. Even when these prerequisites are met, complications beyond the control of the evaluator frequently arise that preclude any definitive conclusions about the project effectiveness. To paraphrase Murphy's Law, "If anything can go wrong in evaluating an educational program, it probably will."

Even in light of the many shortcomings of the evaluation reports we analyzed, we found a great deal of evidence to suggest that the Colorado Bilingual Programs have been effective in improving the English reading achievement of L.D. students. This is particularly true if projects that prevented a deterioration in the performance of their participants are defined as being successful. We would strongly suggest that this phenomena not be lightly dismissed. The fact that we found a number of projects that evidenced very substantial gain at one or more grade levels (see TABLES I-1 and I-4) is even stronger support for the efficacy of the programs. Any program evidencing an average gain of seven or more NCE's has had a substantial impact. Even if the data reported is viewed very skeptically (and we believe that we have been very skeptical indeed) it is hard to avoid concluding that the programs have often been quite successful in meeting this objective of The Act.

TABLE I-4

Pre-Post Test Change -- Students in B/E Programs
L.D. = X; Non-L.D. = O

Proj.	Grade	Lost 7 or more NCE units	Pre & Post Tests Within + 6 NCE units	Gained 7 or more NCE units
#3	K			
	1			XO
	2			
	3			
#5	K			
	1		X	O
	2			
	3			
#6	K			
	1			
	2	O	X	
	3		XO	
#9	K			
	1		XO	
	2	XO		
	3			
#12	K			
	1		X	O
	2	X	O	
	3		XO	
			X	
#13	K	XO		
	1			
	2		XO	
	3		XO	
			XO	
#14	K			
	1			
	2			
	3		XO	
			O	X
#15	K			
	1			XO
	2			
	3		X	O
			X	O
#17	K			
	1			
	2		XO	
	3		XO	
			XO	
#20	K			
	1			
	2		XO	
	3			
#21	K			
	1			XO
	2			
	3			
#23	K			
	1			
	2			XO
	3		O	X
			XO	

TABLE I-4 (cont.)

Pre-Post Test Change -- Students in B/E Programs
L.D. = X; Non-L.D. = O

Proj.	Grade	Lost 7 or more NCE units	Pre & Post Tests Within + 6 NCE units	Gained 7 or more NCE units
#24	K			
	1			
	2		XO	
	3		XO	
#25	K			XO
	1			XO
	2			XO
	3			XO
#26	K			
	1			
	2		X	O
	3			
#27	K			XO
	1			
	2		X	O
	3		O	X
#29	K		X	O
	1			XO
	2			
	3			
#30	K			
	1			X
	2	X		
	3			X
#31	K			
	1			
	2			X
	3			
#33	K			XO
	1			
	2			
	3			
#34	K			
	1			
	2		XO	
	3		XO	
#35	K		O	X
	1			XO
	2		XO	
	3		XO	
#36	K		XO	
	1	X	O	
	2			XO
	3			
#37	K			
	1		XO	
	2		XO	
	3	X		O

II

IMPROVED SCHOOL ATTENTANCE AND A REDUCED DROP OUT RATE

Colorado Bilingual Programs were requested to report the number of possible student attendance days and the number of actual student absence days. The evaluation reporting forms requested that this information be reported for each grade level and separated so as to reflect the attendance of Linguistically Different students in the program as well as Non-Linguistcally Different students in the program. Districts were also requested to report similiar figures for Linguistically Different students that were not participating in the bilingual program and Non-Linguistically Different students not in the bilingual program. Twenty-four of the 33 projects reporting provided at least some of the requested data.

Where projects reported the requested data, DACBE computed attendance ratios for each grade level and student classification for each project. This computation yields a ratio that can be read as the percent of total possible days that students actually attended. If accurate reporting is assumed, then the total possible days a student could have attended would be sensitive to students that enter the school after the beginning of the academic year and/or leave the school prior to the end of the year. Thus total possible days attended is not assumed to be a simple product of the number of students times 180 school days.

This objective of the Act calls for an improvement in attendance and a reduced drop out rate. The terms "improved" and "reduced" imply a comparison to some baseline, or previously established state of affairs. No data of this nature was available. Therefore, determining the success of the Colorado

Bilingual Programs in meeting this objective is impossible based on reported data. It is only possible to describe the reported attendance. Any assumptions or inferences that may be made regarding the effect of the program on the reported attendance are of questionable validity. Since no attendance baseline data, or other estimate of expected attendance rates in the absence of the program, are available, no sound determination of the effect of the program on attendance can be made.

Since the evaluation reporting form did not call for any data relating to the drop out rate, it is impossible to form a judgment on the program's effect on drop out rates. Since the majority of the school drop outs occur at the secondary level, it is unlikely that the bilingual program has any significant impact on the drop out rate of the participants in early elementary school. The drop out rate for kindergarten thru third grade is usually quite small.

Although it is not possible to judge the effect of bilingual programs on the attendance of the program participants, the data reported by the Colorado Bilingual Programs is summarized and presented below.

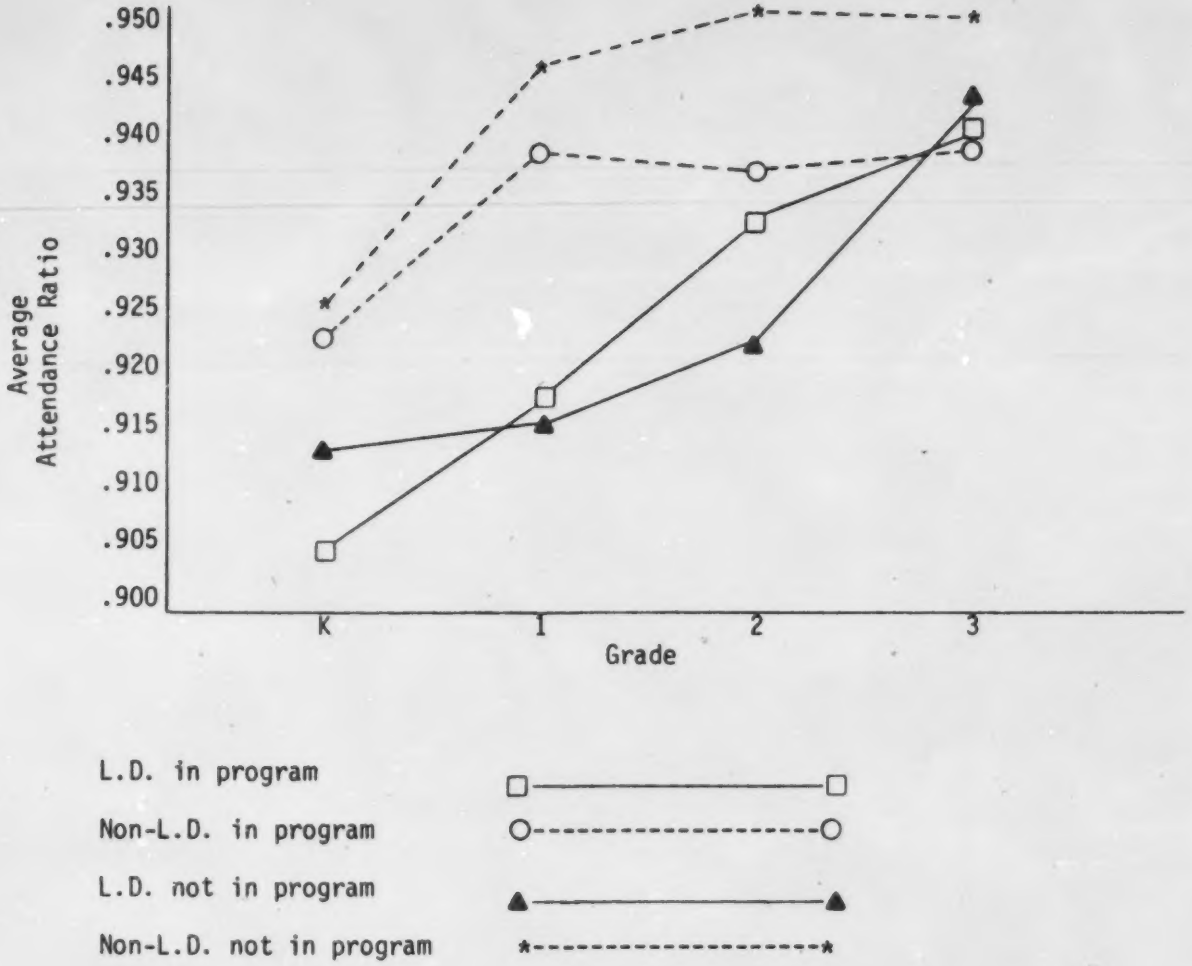
Table II-1 displays the average ratios of actual days attended to total possible days for all of the bilingual programs reporting. On the average, Linguistically Different students participating in the bilingual programs were in attendance at least 90 percent of the time. Linguistically Different kindergarteners in the program attended 90 percent of the time. First graders attended almost 92 percent of the time, second graders over 93 percent of the time and third graders 94 percent of the time. Non-Linguistically Different students in the bilingual programs attended 92 percent of the time in kindergarten and almost 94 percent of the time in first grade. Second and third grade attendance ratios are not significantly different between Linguistically Different and Non-Linguistically Different students participating in the program. Figure II-1 graphically displays the data presented in Table II-1. Although this

TABLE II-1

Statewide Average Ratios Actual Days Attended to Total Possible Days

Grade	In Program		Not in Program		Row Averages
	L.D.	Non-L.D.	L.D.	Non-L.D.	
K	n = 24 .904	n = 23 .923	n = 6 .913	n = 12 .925	n = 65 .915
1	n = 24 .918	n = 23 .938	n = 6 .915	n = 13 .945	n = 66 .930
2	n = 21 .933	n = 20 .936	n = 6 .922	n = 11 .950	n = 58 .936
3	n = 16 .940	n = 16 .939	n = 7 .942	n = 9 .949	n = 48 .942
Column Averages	N = 85 .922	N = 82 .934	N = 25 .924	N = 45 .942	

n = number of projects reporting specified data.



Average attendance ratios of bilingual and non-bilingual education groups, K-3rd grade, L.D. and non-L.D. students (24 projects reporting.)

FIGURE II-1

data is cross sectional rather than longitudinal in nature, it suggest that Linguistically Different and Non-Linguistically Different children begin their school careers with somewhat different attendance patterns. These differences become negligible for these children in the bilingual programs during the second and third grades.

It should be noted that, overall, attendance is strong for all groups of students represented in the reported data. The differences between the Linguistically Different and Non-Different students, both in and out of bilingual programs, that is depicted in the graph, is relatively small. These differences fall within a range of 5 percent of the total possible class days actually attended.

Table II-2 displays the project by project attendance ratios for all grade levels and student groups reported within each project. Examination of this table reveals that virtually all attendance ratios are in excess of .85. There are only three instances where a grade level group of children in an individual project attended less than 85 percent of the possible days.

Several projects reported attendance data that we were unable to use. Five districts reported their attendance in terms of the average number of days attended or the average days absent. Two projects computed attendance ratios based on a standard assumption of 180 potential contact days per student, rather than on the real possible number of student contact days. One project reported the total days attended by all students in each grade. That project did not report the total possible days. Another project simply stated "77 of 123 children had fewer absences in 1977-78 than 1976-77". Two additional projects did not distinguish between grade levels or student classification in reporting their attendance data. One project reported no attendance data whatsoever.

TABLE II-2

Ratios of Actual Days Attended to Total Possible Days

(24 Projects Reporting)

Project	Grade	In Program		Not in Program	
		LD	Non-LD	LD	Non-LD
1	K	.865	.935		.931
	1	.910	.936		.929
	2	.920	.922		.956
	3				
2	K	.938	.956	.876	.932
	1	.945	.950		.959
	2	.956	.952		.983
	3	.956	.942		.936
3	K	.906	.949		
	1	.915	.957		
	2			.935	.964
	3			.953	.954
5	K	.972	.951		.936
	1	.923	.957		.942
	2				
	3				
6	K	.874	.909		
	1	.990	.923		
	2	.928	.926		
	3			.929	.935
12	K	.920	.950		
	1	.946	.961		
	2	.968	.978		
	3	.953	.955		
13	K	.932	.936	.972	.958
	1	.948	.933	.949	.943
	2	.945	.951	.922	.948
	3	.953	.943	.940	.944
14	K	.902	.917	.933	.937
	1	.934	.932	.934	.938
	2	.951	.951	.965	.949
	3	.952	.956	.946	.949
15	K	.899	.875		
	1	.915	.934		
	2	.945	.924		
	3	.916	.931		
16	K	.866	.923		
	1	.850	.926		
	2	.917	.905		
	3				
17	K	.887	.874		
	1	.885	.929		
	2	.899	.902		
	3	.948	.895		
18	K	.873	.909	.924	.864
	1	.908	.934		.941
	2	.924	.934	.961	.953
	3	.908	.932	.955	.953

TABLE II-2 cont.

Project	Grade	In Program		Not in Program	
		LD	Non-LD	LD	Non-LD
21	K	.930	.919		.941
	1	.927	.936		
	2	.952	.942		
	3	.941	.909		
23	K	.885	.922		.856
	1	.867	.911	.800	.939
	2	.889	.870	.850	
	3	.889	.933	.911	
24	K	.941	.939		
	1	.940	.961		
	2	.951	.952		
	3	.966	.964		
25	K	.924	.918	.926	.935
	1	.938	.940	.935	.958
	2	.942	.940	.901	.940
	3	.954	.948	.958	.960
26	K	.909	.914		
	1	.879	.902		
	2	.886	.927		
	3	.927	.935		
27	K	.914	.928		.932
	1	.925	.947		.946
	2	.923	.943		.946
	3	.948	.942		.941
28	K	.930	.937	.848	.957
	1	.913	.953	.949	.956
	2				
	3				
29	K	.890	.919		
	1	.916	.934		
	2	.884	.927		
	3	.950	.938		
30	K	.826	.889		
	1	.907	.941		
	2	.917	.961		
	3	.926	.952		
31	K	.907			.926
	1	.933			.934
	2	.942			.939
	3				
34	K	.914	.905		
	1	.916	.931		.950
	2	.951	.942		.883*
	3	.946	.955		.972

Projects not included in analysis:

#7, #33, #22, #20, - reports average days absent

#35 - reports average days attended based on 180 possible contract days.

#32 - did not use form

#9, 10, 11 - reported total days attended

#4 - does not differentiate grade levels or L.D./non-L.D.

The available data does not justify any conclusion as to the effect of the bilingual program on improving school attendance or reducing the drop out rate. If the data reported is assumed to be accurate (DACBE had no way to validate it) then it is best used as descriptive for the 1977-78 program year. It is possible that this data could be used in subsequent years as a baseline against which to measure additional program impact. This suggestion should be implemented with caution, however, since one school that reported a relatively low attendance ratio for one grade explained that the low ratio was caused by one child who accounted for 53 of the total 104 absences reported. Therefore, any use of this data for baseline comparison purposes in subsequent years must be sensitive to such unusual circumstances. Failure to account for such occurances could lead to spurious conclusions as to the impact of the program.

Although no causal inferences as to the programs' effect on attendance can be made from this data, policy judgments may be possible. It is not possible to empirically determine whether the attendance ratios reported are good or acceptable. Such policy judgments may be facilitated with the reported information.

If this sort of information is to be collected from bilingual projects in subsequent years, several improvements should be made in the current reporting procedures. The nature of the data to be reported should be explicitly noted on the reporting form itself. At this time the instructions for completing this reporting form are contained in a separate document. As noted above, several projects failed to report this information in the requested format. The reporting problems might be simplified by requesting that attendance data be reported in the form of a ratio of total number of days absent to total possible student class days. Finally, some system should be devised to allow schools to note

unusual absence conditions, such as long term illness or wide spread occurance of contagious disease (such as a flu epidemic). Failure to account for such conditions would result in spuriously high absence ratios being reported.

Summary

Analysis of reported attendance data indicates that students in bilingual programs are in attendance over 85 percent of the possible school days. Average attendance ratios for Linguistically Different students in bilingual programs improves from 90 percent in kindergarten to 94 percent in third grade. Although there is an apparent trend towards improved attendance across the grades, this data does not support either the conclusion that such a trend would be found in a longitudinal, rather than a cross-sectional study, or the conclusion that the bilingual program is responsible for the observed attendance. Since no baseline attendance is available, it is not possible to judge the effect of the program on improving school attendance. Although it might be possible to use this year's data to establish a baseline for individual districts, against which next year's attendance performance could be compared, such a course of action should be approached with caution. Even if improved attendance were noted, establishing a causal relationship between the bilingual program and the improved attendance rate would be difficult. Additionally, since such a comparison would be between consecutive years in which the program was operating, such a comparison would not answer the question "is attendance with the bilingual program better than would be expected in the absence of the bilingual program". This is the crucial question that should be answered in order to meaningfully evaluate the impact of the Colorado Bilingual Programs in meeting the legislative intent.

An attractive alternative would be to formulate a policy goal in terms of a desired attendance rate and then request programs to measure the extent to which their attendance rates meet or exceed the established goal. The present data may provide guidance in establishing feasible attendance goals. Specifying a desirable attendance goal is beyond the realm of this report.

III

"DEVELOPMENT OF A POSITIVE SELF CONCEPT AND ATTITUDE"

This objective is probably the most difficult to accurately measure and to meaningfully interpret. The extensive literature on self concept contains literally dozens of formulations of the construct "self concept". Since there is little agreement on what self concept is, there is little agreement about what ought to be measured in its name. As a result, no two self concept instruments appear to measure precisely the same thing.

Since the legislature has included this as an objective of the bilingual education act, it is clear that the programs sponsored under the act are intended to do something about the self concept of the participating children. This, of course, presupposes that there is a need to do something about these children's self concept: that they enter the program with a less than positive self concept which needs to be improved to a positive self concept. The data contained in the evaluation reports we reviewed provides no evidence to indicate that such a need for improvement exists.

Even if such a need were established, or is assumed, adequate measurement technology to measure the program's impact is virtually nonexistant. Neither commercially published nor locally developed instruments are sufficiently reliable and valid to accurately document program impact on self concept. As a result, it is extremely difficult (if not impossible) to show a causal relationship between an educational program and the self concept of its participants. Any sound attempt to establish such a relationship would require the expenditure of substantial fiscal as well as human resources. The self concept data reported by Colorado projects must be viewed in light of these constraints.

Most commercially developed self-concept measures for early elementary school children attempt to measure various aspects of socialization to the school environment as well as factors more conceptually related to self-concept. The Self Observation Scale (used by two Colorado projects) includes subscales that purport to measure self acceptance, self security, social maturity, and school affiliation. The Primary Self-Concept Inventory (used by nine Colorado projects) purports to assess aspects of self-concept relevant to school success: personal, social, and intellectual domains in self-concept. There is, however, no evidence of any relationship between the score a child obtains on this measure and that child's success in school. According to the publishers, this screening inventory is designed to identify students with undesirably low self-concepts.

Eight of the Colorado projects made use of locally developed measures of self-concept. Most of these appear to be teacher judgment ratings rather than ratings obtained from student responses. The descriptions provided of these locally developed instruments suggest that they are primarly focused at measuring the extent to which children have become adapted to the school environment.

Because of the divergence of opinion on the appropriate meaning of self-concept, and the related variety of factors that self-concept instruments attempt to measure, it is difficult to meaningfully aggregate the data reported by the Colorado bilingual programs. Of the 33 project evaluation reports reviewed, 11 (one-third) reported no data on student self-concept at all. Of the 22 projects that reported some self-concept data, 8 made use of locally developed instruments, 9 used the Primary Self-Concept Inventory, 2 used the Self-Observation Scale, 1 used the California Test of Personality and another used the Piers-Harris inventory. One project went so far as to have a sample of the Linguistically Different students examined by a school psychologist. The psychologist found these students' self-concept to be satisfactory. Another project had members

of a parent group rate the self-concept of their children. These parents perceived their children's self-concept to be adaquately positive. Students in this project were also administered the Primary Self-Concept Inventory.

The data provided suggest that the students in bilingual programs funded by the State of Colorado generally possess adequately positive self-concepts and attitudes. Although several programs reported the results of multiple administrations of self-concept measures (pre-post testing) no reliable estimate of the effect of the program on improving the self-concept of participating students can be made. This state of affairs is as much result of the lack of good instrumentation as it is a result of lack of agreement on what constitutes self-concept.

It is generally true that personality measures of very young children (early elementary school) tend to be unreliable. That is, the measures of a personality trait such as self-concept tend to fluctuate and be unstable from testing to testing. This is due in large part to the fact that children in this age group are developing very rapidly. As a result, little remains static over relatively short time intervals. As a result, when improvement in measured self-concept is reported, it is difficult to determine whether a real improvement, attributable to any educational program, has taken place or whether the child has simply matured between testings. This phenomenon would be particularly expected where the measures tap the extent to which students have become adjusted to the school environment. It is important to also remember that improving self-concept (at least within some theoretical frameworks) is only appropriate if a deficit exists initially. There is no evidence to suggest that students in Colorado bilingual programs suffer from such a deficit, as a group.

Summary

The data reported by Colorado bilingual programs suggests that students in these programs generally have adequately developed positive self-concepts and attitudes. No conclusions can be made on the effect of the program on the students' self-concepts for a number of reasons, both theoretical and psychometric. In light of these findings, we recommend that less effort be spent in attempting to evaluate the impact of the bilingual programs on the self-concept of their participants. We feel that the fiscal and human resources currently allocated to evaluating this aspect of the program could be more profitably allocated to evaluating the academic impact of the program.

GREATER PARENTAL INVOLVEMENT IN THE SCHOOL PROGRAM

Assessing the effectiveness of the Colorado Bilingual Programs in meeting this objective of the act requires a comparison of current parent involvement levels with previous (pre-program) levels of parent involvement. Unfortunately, no such base line for comparison exists. As a result, it is impossible to judge the extent to which the programs have resulted in greater parent involvement. The current evaluation reporting forms do not request that such information be reported. Even if they did, it is doubtful that most school districts would be able to provide valid or reliable estimates of pre-program parent involvement levels.

Although it is not possible to draw inferences about the project's effect in improving parent participation, it is possible to describe the reported levels of parent participation. Projects are requested to complete two reporting forms for the Colorado Bilingual Education Unit. The first of these forms is intended to describe the parent advisory committee operating in each program. This form asks districts to report the make-up of the parent advisory committee (PAC) in terms of total membership, total members who are parents of linguistically different students, total regularly scheduled meetings and total special meetings. This reporting form also requests an estimate of the extent to which the PAC was involved in each of 17 different activities. The specified activities correspond to activities mentioned in the act.

Table IV-1 summarizes the information reported regarding the membership of PACs. The average PAC consists of 13 members, although some PACs have as many as 30 or as few as 4 members. Fifty-four percent

TABLE IV-1

PAC Membership*

(1)	Total Members	429	Avg. per committee	13.0	Range 4-30
(2)	Total members who are parents of L.D. students	231	Avg. per committee	7.0	Range 0-20
	% of total PAC membership represented by L.D. parents	54%	# of committees with 75% (or better) L.D. representation (included one 73%)	11	
	% range	0-100%	# of committees with less than 75% L.D. representation (11 below 50%)	22	
(3)	Total regularly scheduled meetings	295	Avg. per committee	9	Range 0-48
(4)	Total special meetings	91	Avg. per committee	3	Range 0-13

* 33 PAC's reporting data; 2 projects did not report this data.

of the average PAC is comprised of parents of Linguistically Different students. Although the average committee has seven such parents in its membership, one PAC has no parents of Linguistically Different children among its membership.

On the average, these committees met nine times during the school year (once a month). One committee reported 48 regular meetings (one every week during the school year). Another committee reported no meetings during the school year. In addition to the regularly scheduled meetings, the average committee met three times for special meetings.

Table IV-2 displays the number and percent of parent advisory committees that reported involvement in each of the 17 specified activities. The reporting form requested that the respondent not only indicate whether the PAC had been involved in each of the listed activities, but also classify the extent of that involvement in each of the activities. The distribution of the responses reflecting extent of involvement is included in Table IV-2. Since a large proportion of the activities listed on the reporting form do not readily lend themselves to the "extent of involvement" scale, this discussion will focus primarily on whether the committees indicated that they did or did not participate in the enumerated activities.

Examination of Table IV-2 reveals that sixty percent of the PACs did not participate in interviewing aide applicants and forty-eight percent of these councils indicated that they neither received aide applications nor made recommendations regarding classroom materials. Table IV-3 lists the activities that parent advisory councils most frequently reported not participating in.

TABLE IV-2

Number and percent of PAC's reporting involvement in specified activities.*

	Extent of Involvement							
	All of the Time		Most of the Time		Some of the Time		Never	
	N	%	N	%	N	%	N	%
(5) Reviewed current evaluation reports**	16	46	6	17	6	17	6	17
(6) Reviewed previous evaluation reports	13	37	8	23	8	23	6	17
(7) Made presentations to the School Board	3	8	4	11	15	43	13	37
(8) Informed other parents about the bilingual education	19	54	11	31	2	6	3	8
(9) Developed new releases about the bilingual program	6	17	6	17	12	34	11	31
(10) Reviewed aide applications	10	28	2	6	6	17	17	48
(11) Interviewed prospective aides	9	26	2	6	3	8	21	60
(12) Made hiring recommendations on aides	13	37	1	2	6	17	15	43
(13) Reviewed the program proposal	22	63	5	14	4	11	4	11
(14) Had input in the program proposal	16	46	8	23	6	17	5	14
(15) Reviewed instructional materials	7	20	3	8	16	46	9	26
(16) Made recommendations on classroom materials	4	11	1	2	13	37	17	48
(17) Participated in teacher/aide inservice	7	20	3	8	13	37	12	34
(18) Helped respond to complaints about bilingual education	12	34	10	28	10	28	3	8
(19) Reviewed the district long range plans for bilingual education	15	43	6	17	5	14	9	26
(20) Received copies of the Bilingual Act	24	68	6	17	3	8	2	6
(21) Received copies of the Rules and Regulations	23	66	6	17	4	11	2	6
(22) Received information concerning the educational needs of children**	12	34	10	28	6	17	6	17

* 35 PAC's reporting data
** 1 PAC did not respond to item

TABLE IV-3

Activities PACs Most Frequently Reported
Never Participating In

# PAC's Never Participating	Activity
13	(7) making presentations to the school board
11	(9) developing new releases about bilingual education
17	(10) reviewing aide applications
21	(11) interviewing prospective aides
15	(12) making recommendations for hiring of aides
9	(15) reviewing instructional materials
17	(16) making recommendations regarding classroom materials
12	(17) participating in teacher/aide inservice
9	(19) reviewing district language B/E plans.

At least 29 (83%) of the PACs report that they participated, at least some of the time, in all of the activities listed in Table IV-4.

TABLE IV-4

Activities PACs Most Frequently Participated In

# PACs Participating	Activity
29	(5) reviewing current evaluation reports
29	(6) reviewing previous evaluation reports
32	(8) informing other parents about B/E
31	(13) reviewing the program proposal
30	(14) giving input to the program proposal
32	(18) responding to the compaints about B/E
33	(20) recieving copies of the Bilingual Act
33	(21) recieving copies of the Rules and Regs.
29	(22) recieving information concerning the educational needs of children

In addition to the activities specified on the reporting form, a number of PACs reported involvement in other activities. These activities are listed in Table IV-5.

TABLE IV-5

Other Reported PAC Activities

- attending bilingual/bicultural seminars and conferences outside the district
- acting as guides or hosts and hostesses for visitors
- planning special parent-student events
- conducting parent-involvement workshops
- reviewing budget reports
- reviewing program progress reports
- making home visits to parents
- receiving teacher applications
- making teacher hiring recommendations
- assisting in revising district census data
- making recommendations to school boards and JBC's

Only one project director was unable to organize a viable PAC, since council members failed to attend meetings even though free transportation was provided. One PAC indicated that their efforts to make significant contributions to their bilingual program and to become involved in relevant activities were repeatedly frustrated by the school administration. In this case, PAC members clearly state that the fact that their council was never involved in most of the specified activities was not due to any unwillingness on the part of the members to become involved.

The information obtained from the reporting forms discussed here should be interpreted with caution. In several instances there was reason to believe that the person that had filled out the form was not reporting activities of the advisory council, but was rather reporting activities engaged in by the project director. In this case, the project director was the individual who completed the form. In other cases the activities reported may not reliably reflect what really transpired. The primary reason for this caution is that the form is unclear as to who should fill it out: the project director or the chairman of the advisory council. One suspects, therefore, that in at

least some cases, responses represent those activities in which the project director was involved, and may or may not be representative of any involvement on the part of the PAC. Furthermore, even when the person preparing the form is identified as the PAC, chairperson, it is difficult to assess the degree to which the entire PAC was involved in the activities. This uncertainty arises because the instructions to the form specifically state, "Please check the column that represents the extent of your involvement in each of the following activities."

In addition to information about the activities of the PAC, projects were asked to report what other parent activity occurred in the program. The reporting forms for this data requested that each parent involvement activity be listed separately, along with the number of parents attending, and the number of LD and Non-LD students that these parents represented. We suspect that the attendance data, and the student representation data, frequently represent estimates rather than accurate records. It was not unusual to find that a project had reported that the number of parents participating in a specified activity corresponded to exactly one (or two) parents per Linguistically Different child in the program. Although it is certainly possible that every child in a program would be represented at a parent activity by exactly one or two parents, it is more likely that the parents of at least a few of the program children would be unable to attend in a given parent activity. Several reports stated that the figures provided were estimates. We suspect that several other projects also estimated these figures. The fact that a large majority of the reported data show a perfect one-to-one correspondence between the number of parents in attendance at any given activity and the number of children in the bilingual program, lends support to our suspicion.

Even when the schools based their attendance figures on actual records, there are problems in interpreting the information they have provided. For instance, one school was careful to count each parent only once for functions such as open houses or classroom visits or parent-teacher conferences. Other schools simply recorded the total number of parents that attended such functions. Thus, it is possible that a school may report 60 parents visiting classrooms when, in reality, 20 parents visited classrooms on three occasions. The actual parent participation would, therefore, be one third of what is indicated by the reported attendance for classroom visits.

Since the reporting forms are open-ended, a great deal of variation occurred in the type of activities that were reported. Some districts provided extensive lists of activities that were highly detailed. For example, "Kindergarten play, September 14, five parents attending". Other districts reported much more general information such as "regular parent activities, 2,832 attending."

The final difficulty in drawing any conclusions from the data provided is that school districts reported activities that they considered to be parent involvement activities. In at least some cases the activities reported would probably not be generally thought of as parent involvement (for example, teacher home visits). Since some things were included as parent involvement activities that probably weren't, it is reasonable to assume that some activities that did occur were not reported. For instance, at least one project listed a workshop for parents on the inservice training reporting form rather than the parent participation reporting form.

TABLE IV-6

Number and Percent of Projects Reporting Various Types of Parent Participation*

Type of Activity	No. of schools reporting activity	% of all projects
I. Regular parent activities		
a. parent-teacher conferences	17	48
b. classroom visits	14	40
c. open house	6	17
d. all of above	2	5
Total reporting regular functions	22	63
II. Volunteer activities		
a. classroom aides	13	37
b. materials development	1	?
c. field trips	10	28
d. parent classroom presentation	12	34
e. all of above	4	11
Total reporting volunteer activities	23	66
III. Athletic events	3	8
IV. Social activities		
a. Traditional holidays	9	26
b. Cultural celebrations	8	23
c. Special programs/parties	17	48
d. Picnics/Teas/Coffees/Luncheons/Dinners	10	28
e. all of above	6	17
Total reporting social activities	27	77
V. Fund raising activities		
a. book fairs	2	5
b. fiestas	1	2
c. community fun day	1	2
d. miscellaneous	4	11
Total reporting fund raising activities	6	17
VI. Parent education	9	26
VII. Bilingual/Bicultural committee programs and projects	9	26

* 35 projects reporting data

As a result of these concerns we felt that the most appropriate treatment of the data provided would consist of simply summarizing the types of parent participation that were reported, and the number of projects that reported each type of participation. Accordingly we made a list of all parent activities that were reported by the projects. These activities were then sorted into general categories. The number of projects that reported such activities were then tabulated. The results of this process are displayed in Table IV-6. No attempt was made to determine the number of parents that participated in these activities or the number of children that these parents represented.

Seventy-seven percent of the projects reported parent participation in a variety of social activities ranging from cultural celebrations to parties, picnics and luncheons. Sixty-six percent (23) of the projects reported that parents participated in volunteer activities such as making a classroom presentation, serving as an aide in the classroom, or going on field trips. Twenty-two of the projects reported that parents engaged in regular parent activities such as parent-teacher conferences, classroom visits and open houses. Nine project reported that parents were involved in some form of parent education program. Six projects reported that parents engaged in various fund-raising activities.

Summary

The thirty-three PACs reporting include 429 members, 321 (54%) of whom are parents of Linguistically Different children. There is considerable variation in the composition of individual PACs. One council reported no parents of LD students among their members and three councils were composed entirely of parents of LD children. Parents of Linguistically Different students constitute at least half the membership of twenty-two of the thirty-three PACs reporting membership data.

The average PAC held approximately nine regular meetings during the reporting period. With the exception of one council that did not meet at all and one council that met 48 times, the remaining 31 PACs held from three to twelve regularly scheduled meetings.

PACs are most likely to be involved in activities such as: reviewing evaluation reports and program proposals, giving input to program proposals, giving bilingual education information to other parents and responding to complaints about bilingual programs, receiving copies of the legislation and rules, and receiving information regarding the needs of participating children.

PACs are least likely to have been involved in: interviewing prospective aides, reviewing aide applications, making recommendations for hiring aides, making presentations to the school board, participating in teacher inservice, reviewing district bilingual education plans and reviewing instructional materials.

We suggest that subsequent data collection activities on PAC membership and activities utilize a modified reporting form that specifies more clearly who is to report the requested data, and whose activities the data should reflect. We also suggest that no attempt to estimate the extent of involvement in each of the listed activities be made. This is primarily because many (if not most) of the listed activities cannot be meaningfully described by whether a PAC engages in them "all of the time", "most of the time", "some of the time", or "never". We think that a simple "yes, no" format will yield as much information.

The majority of projects report parent participation in regular parent activities, volunteer services and social functions. Whether the current levels of parent participation represent an increase over past

levels cannot be determined for two reasons: 1) baseline data concerning past levels of parent involvement are not readily available, and 2) there are many discrepancies in reporting styles, data collection methods and, we suspect, a substantial amount of unreported parent activity.

The open-ended nature of the reporting form provides virtually no guidance for what sorts of activities should be reported under the auspices of parent participation. If a shopping list of parent activities, such as the one presented in Table IV-6, were presented to all Colorado projects with the request that they indicate which activities had occurred in their projects, we suspect that the overall incidence of many of the listed activities would appear to be much greater.

V

CONCLUSIONS AND RECOMMENDATIONS

IMPROVED PERFORMANCE IN COMPREHENSION, READING, WRITING, AND SPEAKING THE ENGLISH LANGUAGE.

Although much of the data that was reported could not be used for this analysis, we were able to identify seven programs out of thirteen in which the L.D. Kindergarten children evidenced significant gains from pre- to post-testing. Five of eleven programs showed significant improvement for first grade L.D. students. Four of eighteen programs showed substantial improvement for second grade L.D. students and four of fourteen programs showed substantial improvement for third grade students. Programs evidencing this level of gain have shown a great deal of impact on the students' reading ability in English. Although not showing this level of gain, many were successful in maintaining the pre-test performance status of participating L.D. students at post-testing. Five of thirteen projects reported no change in pre-post-test status for Kindergarten L.D. students. Three of eleven first grade projects showed no change for L.D. students between pre- and post-testing. Thirteen of eighteen second grade programs showed no change from pre- to post-testing for L.D. students and nine of fourteen projects reported no change from pre- to post-testing for third grade students. Although these programs did not meet our criteria for significant gain, they were successful in that they were able to prevent the participating students from experiencing the loss in relative reading achievement that is frequently observed in Mexican-American students in public schools.

In addition to these findings, the Bilingual Unit of the Colorado Department of Education has informed us that, according to a survey they recently conducted, approximately 15% of the students that were classified as

Linguistically Different during the 1977-78 school year showed sufficient gains at the end of the year to be reclassified, in accordance with the provisions of The Act, as Non-Linguistically Different students. That is, fifteen percent of the L.D. students in the program during 1977-78 reached or exceeded the district average in English reading achievement.

Although not all districts were successful in improving the English reading skills of their participating students, we believe that many of the bilingual programs are experiencing high levels of success.

We are concerned that there appears to be great room for improvement in the evaluation of the State programs. We feel that each program should allocate adequate fiscal resources to its own evaluation to yield valid and reliable documentation of its effect. Good evaluations usually require that between six and twelve percent of the total operating budget of the project be allocated to evaluation. Small programs must spend proportionately more than large ones. Projects should exercise caution in employing evaluators. Few who present themselves as qualified really are.

At a minimum, we would suggest that projects report the average pre-test score and average post-test score for the Linguistically Different and Non-Linguistically Different children at each grade level within their project accompanied by the N and standard deviation for each of those means. These statistics should include only those children who took both the pre- and the post-test. The mean, standard deviation and N for the L.D. and Non-L.D. students (reported separately) who took only the pre- or only the post-test should be reported as well. It would also be helpful if the correlation between the pre- and the post-test for each group of students at each grade level was reported. All scores should be reported in terms of NCE's. Percentiles, raw scores and grade equivalent scores are of virtually no use.

Making sense out of scores like this requires a great deal of time and effort. Projects should make us of standardized, norm-referenced achievement tests. When selecting such tests, they should be sure that any test they decide to employ has norms based on real data, corresponding to the dates in the school year that they plan to conduct their testing. Some tests should never be used for Fall to Spring testing. After selecting an achievement test, the district should be sure to administer the test within plus or minus two weeks of the date that the test publishers collected the normative data. Every effort should be made to obtain test scores for all children in the program, even if this means giving several make-up tests within a week of the main testing.

IMPROVED SCHOOL ATTENDANCE AND REDUCED DROP-OUT RATE

We were unable to make a determination as to whether the attendance rates reported represent an improvement over pre-program levels since no such base line data was reported. Students in bilingual programs are, however, in attendance at least 85% of the possible school days. The average attendance for Linguistically Different students in bilingual programs improves from 90% in Kindergarten to 94% in third grade. Non-Linguistically Different students in the bilingual program show slightly higher attendance figures for Kindergarten and first grade than do L.D. students, although there is no apparent difference between the attendance rate of L.D. and Non-L.D. students in the bilingual programs in second and third grades. Since no data regarding the drop-out rate of students in the programs was reported, we are unable to offer any statement as to the effect of the programs on reducing the drop-out rate. Since the drop-out rate of students in Kindergarten through third grade is negligible, we suspect that the program has little effect on it.

We would recommend that a policy decision be made specifying a goal for attendance. We would then recommend that projects report their success in

meeting this goal. Since pre-program attendance rates are unlikely to be readily available, we doubt that it is possibe to determine the effect of the program on improving school attendance.

DEVELOPMENT OF A POSITIVE SELF CONCEPT AND ATTITUDE

The data reported by the Colorado Bilingual Programs suggest that students in these programs generally have adequately developed positive self concepts and attitudes. No conclusions can be made of the effect of the program on the students' self concepts for a number of reasons, both theoretic and psychometric. As a group, these students do not appear to need to improve in these areas.

In light of these findings, we recommend that little effort be spent attempting to evaluate the impact of the bilingual programs on the self concept of their participants. We feel that the fiscal and human resources that are currently allocated to evaluating this aspect of the program would be much more profitably allocated to evaluating the academic impact of the program. If any attempt is made to evaluate this aspect of the program, we strongly urge that it be anecdotal in nature. The information obtained from most currently available self concept measures is hardly worth the price of the test.

GREATER PARENTAL INVOLVEMENT IN THE SCHOOL PROGRAMS

Two sorts of data regarding parent involvement were reported. The first pertained to the activities of the parent advisory council (PAC). The second pertained to the general parent activities in the school system. Since no pre-program base lines of parent involvement levels were available, it was not possible to make inferences about the program's effect on improving parent participation. We were only able to summarize and describe the activities reported.

Only one program reported that a PAC was not formed. The average PAC held nine regular meetings during the school year. One council did not meet at all and one council met 48 times. PACs were most likely to have been involved in activities such as: reviewing evaluation reports and program proposals, giving input to the program proposals, giving bilingual education information to other parents and receiving copies of the legislation and rules and receiving information the needs of participating children. PACs were least likely to be involved in activities such as: interviewing prospective aides, reviewing aide applications, making recommendations for hiring aides, making presentations to the school board, participating in teacher inservice, reviewing bilingual educational plans and reviewing instructional materials.

We suggest that subsequent data collection activities on PAC membership and activities utilize a modified reporting form that specifies more clearly who is to report the requested data and whose activities the data should reflect. We would also suggest that the forms simply ask whether the PAC had participated in the specified activities.

The majority of projects report parent participation in a number of activities within the school program. The most frequently reported parent activities are volunteer services and social functions. Whether the current levels of parent participation represent an increase over past levels can not be determined for several reasons. First, base line data concerning the previous levels of parent involvement are not available. Second, the current reporting forms are open-ended, and therefore little uniformity between reports was encountered. We would recommend that instead of an open-ended form asking the project to list all of the parent activities, that a "shopping list" of possible parent activities be generated and projects asked to report, on a yes or no basis,whether they had experienced such activities. If more

detailed information is desired, projects might be asked to report how many parents attended the activities and how frequently each activity occurred. We are not in strong support of such a reporting procedure since we believe that its accuracy is easily compromised. Providing accurate information on the number of parents participating in each activity might well require more time and effort than the resulting information would be worth. This is an area in which anecdotal reports would be useful.

APPENDIX A

School Districts Submitting Evaluation Reports

Adams County Sch. Dist. #50

Arkansas Valley Brd. of Co-op Ed. Serv.

Colorado Sprgs. Pub. Schools

Durango 9-R

Huerfano School District RE-1

Johnstown RE-5J

Lamar, RE-2

Los Animas Pub. Sch. Dist. RE-1

Mapleton Public Schools

Mesa County Valley Sch. Dist. #51

Montezuma - Cortez Sch. Dist. RE-1

Pagosa Springs

Pueblo Sch. Dist. #60

Southwest Board of Coop. Serv.

- Dolores
- Egnar
- Mancos

St. Vrain Valley RE-IJ

Weld County Re-2 (Eaton)

Weld County Re-6 (Greeley)

Weld County RE-8 (Fort Lupton)

Weld County RE-9

Adams County - Dist. 14

Harrison School Dist. #2

Ignacio Sch. Dist. 11

Montrose Sch. Dist.

Boulder School Dist. RE-2

Eagle County Sch. Dist RE-50J

Delta County Sch. Dist. 50

27J Brighton

Monte Vista C-8

RE7 Platte Valley, Kersey

Denver Pub. Schools #1

Lake County School District. R-1

Prowers County Re-1

San Luis Valley BOCES

South Central BOCES

- Trundad Dist. 1
- Pueblo #70
- Primero R-2

South Conejos RE-10

REFERENCES

Coleman, J. S. et al, Equality of educational opportunity. Washington D.C.: U.S. Government Printing Office.

Horst, D.P., Tallmadge, G.K. & Wood, C.T. A Practical guide for measuring project impact on student achievement. Washington, D.C.: U.S. Government Printing Office, (1975)

Tallmadge, G.K. & Horst, D.P. A procedural guide for validating achievement gains in educational projects. Washington, D.C.: U.S. Government Printing Office, (1976)

NATIONAL DISSEMINATION AND ASSESSMENT CENTER FOR BILINGUAL EDUCATION

EXECUTIVE SUMMARY

Analysis of the final Evaluation Reports submitted to the Bilingual Unit of the Colorado Department of Education by bilingual education programs funded under the Colorado Bilingual/Bicultural Education Act during the 1977-1978 school year.

by

Ross P. Goldsmith
Assessment and Evaluation Coordinator

and

Marilyn Babcock
Evaluation Intern

Assessment Staff:

Ross P. Goldsmith
Evangelina Mangino
Sheila Reifle
Marilyn Babcock

7703 NORTH LAMAR AUSTIN, TEXAS 78752 TELEPHONE: 458-9131

BACKGROUND:

At the request of the Bilingual Education Unit of the Colorado Department of Education, the Assessment Component of the Dissemination and Assessment Center for Bilingual Education at Austin, Texas agreed to review and synthesize the evaluation reports submitted to the Bilingual Unit by the programs operating under the Colorado Bilingual Bicultural Education Act during the 1977-78 school year. We undertook the role of external evaluator as a service to the Colorado Department of Education.

The primary focus of our activities was to analyze and present the data contained in the evaluation reports and to determine the extent to which the programs have been successful in meeting the objectives of The Act. The focus of our efforts has been to determine the success of the Colorado Bilingual Education Programs in:

- Improving performance in Comprehension, Reading Writing and Speaking the English Language.
- Improving school attendance and reducing the drop-out rate,
- Developing positive self concepts and attitudes, and
- Increasing parental involvement in the school program.

We also provided suggestions for improving subsequent evaluation efforts. We did not address such process variables as how well the projects met their own objectives, how well they had assessed the needs of participating children, or how closely the programs followed prescribed instructional strategies. Likewise, we did not address any effects the programs may have had that are not directly related to the objectives of The Act. Our efforts were directed toward determining how much better students did in the bilingual projects than they would have done without them.

METHODOLOGY:

The general design of the study was to compare how well students did in the program compared to how well they would have been expected to do without it. Since no control groups were available, we made use of a norm referenced evaluation design using the pre-test scores of participating students to generate an expected post-test score against which to compare the observed post test score of the participants. By necessity, we used the mean pre and post test score of students within each grade level (K-3) and student classification (Linguistically Different and Non-Linguistically Different) of each project reporting pre and post test scores from standardized, norm referenced, achievement tests. Since insufficient data was available to perform significance tests, we established a criteria by which to judge whether the observed pre-post test differences were educationally significant. Our criterion for significant improvement (or loss) was 1/3 of a standard deviation (7 Normal Curve Equivalents).

Since no comparison baseline was available to determine whether attendance or parent participation had improved over previous levels, we simply summarized and reported the attendance and parent involvement levels reported by the participating programs. We also briefly summarized and discussed the data presented relating to the self concepts of the participating students.

The first step in our analysis was to read each evaluation report thoroughly. The statistical data pertinent to each objective of The Act was then summarized and spot checked to guard against mathematical and transcription errors. After the data was summarized, it was reviewed and interpreted. The discussion focused on interpreting the data, specifying the limitations of the finds and making recommendations for improving subsequent data collection activities. The evaluation reports reviewed for this report represent 40 school districts across the state implementing bilingual education programs in one or more grades K-3.

FINDINGS:

Improved performance in comprehension, reading, writing and speaking the English language.

The data provided addressed achievement in English reading as asssed by several well recognized, nationally normed reading tests. We define a very successful program as one in which students gained an average of 7 Normal Curve Equivalent (NCEs) in reading achievement. We argue that programs in which Linguistically Different (LD) students did not experience substantial losses should be viewed as successful since there is a large body of literature suggesting that LD students normally fall further and further behind throughout public school. Therefore, we feel that programs which prevent the expected decline in achievement standing are at least moderately successful. Unsuccessful programs are defined as those in which LD students demonstrated an average loss of 7 or more NCEs.

Not all projects reported pre and post test scores for the children in each grade in the bilingual programs. Of the thirteen that reported this data for Kindergarten LD children, seven evidenced significant gain, 5 evidenced no change and one reported a loss. Of the ~~eleven~~ reporting this data for First grade LD students, 5 reported a gain, 3 held their own and 3 lost ground. Of the eighteen reporting for Second grade LD children, 4 gained, 13 held their own and one lost ground. In Third grade, fourteen programs reported data for their LD pupils. Four of these reported a significant gain, 9 reported no gain or loss and one reported a loss. In addition to these findings, the Bilingual Unit conducted a survey and found that over 25% of the children that were classified as LD during the 77-78 school year were subsequently reclassified as Non-Linguistically Different (NLD) at the end of the year because they had reached or exceeded the district average in achievement scores.

Close examination of the achievement data for NLD students in the bilingual programs reveals that their average post test scores were almost always at or above the 50th percentile of the nationally normed tests. Although we can not determine whether the bilingual program was primarily responsible for their performance, we feel that they have not been harmed by their participation in the programs. No project reported average NLD post test scores at or below the 37th percentile.

Although there is great room for improving the quality of the evaluations, the data submitted support the conclusion that the bilingual programs funded under The Act have been generally effective in improving the English reading skills of participating students in the 24 programs reporting some useable pre and post test data. This is particularly true if projects that prevented a deterioration in the performance of their participating LD students are defined as successful. The fact that we found thirteen programs

that evidenced very substantial gain (7 NCEs) at one or more grade levels is even stronger support for the efficacy of the programs. Any program resulting in an average gain of seven or more NCEs has had substantial impact.

Twenty-four projects reported pre-post data for LD students in at least one grade. This data represents 56 grade levels within these projects. Of these 56 grade levels within projects, 20 (36%) showed educationally significant gains in English reading achievement, 30 (54%) were successful in preventing losses and only 6 (10%) were unsuccessful, showing significant losses in English reading achievement. In 6 (10%) of these grade levels LD students whose average scores fell below the 37th percentile on the pre-test scored above the 50th percentile at post-testing.

Improved school attendance and a reduced drop out rate

Since no base line attendance data was available, we were unable to determine whether reported attendance levels represent an improvement over past levels or are due to the bilingual programs. We were, however, able to summarize the attendance data reported.

Overall, LD students in bilingual programs attended school between 90 and 95 percent of the possible days during the 1977-78 school year. Kindergarten LD students attended an average of 90% of possible days. First grade LD students attended an average of 92% of the possible days. Second grade LD students attended 93% of the possible days and Third graders attended an average of 94% of the possible days. No project reported that the LD students in any one grade attended less than 85% of the time. Several reported attendance rates for LD students in at least one grade at or above 95%.

Non-Linguistically Different students evidenced slightly higher attendance in Kindergarten and First grade than did LD students (92% and 94% respectively). In Second and Third grade, NLD students' attendance rates were virtually identical to those of LD students (94% for both grades). We doubt that these attendance rates can be improved upon significantly.

No data pertaining to the drop out rate of children in the bilingual programs was provided. Therefore we are unable to determine the effect of the programs on reducing it. However, since the drop out rate for children in Kindergarten through Third grade is usually negligable, we doubt that a bilingual program could have much effect on it.

Development of a positive self concept and attitude

This objective is the most difficult to accurately measure and to meaningfully interpret. There is little agreement about what self-concept is or how to measure it. Available measurement instruments are technically weak and do not measure the same things. As a result, we were unable to meaningfully aggregate the reported data or offer conclusions on the effect of the program on the students' self concepts and attitudes. The data presented suggests that the students in bilingual programs funded under The Act have, as a group, adequately positive self concepts and attitudes. The data does not suggest that these students, as a group, need to improve in these areas. We recommend that evaluating this component of the program be a low priority.

Greater parental involvement in the school programs

Two types of data regarding parent involvement were reported. The first pertains to the activities of the Parent Advisory Council (PAC). The second reflects general parent involvement activities. Since no pre-program baseline of parent involvement levels were reported, we were unable to determine whether the reported levels represent an increase, a decrease or no change from previous parent involvement.

Only one bilingual program was unable to form a PAC, although the administration in that district made several efforts to do so. The remaining projects were able to form PACs that met an average of once a month. PACs were most frequently involved in activities such as: reviewing proposals and evaluation reports, providing input to proposals, providing information to parents and reviewing the legislation and rules. PACs were least likely to participate in: the process and decision making of hiring aides, making presentations to the school board, teacher inservice, reviewing bilingual educational plans and reviewing materials.

A wide variety of parent activities were reported. These ranged from volunteer activities such as serving as aide in the classroom to fund raising activities. The most frequently reported activities were volunteer services, social functions and regular parent activities such as teacher conferences.

LIMITATIONS OF THE STUDY

In the body of the report, we have devoted a good deal of space to discussing the limitations of the data reported and the conclusions that can be drawn from it. We have also provided many recommendations for improving the data collection procedures for subsequent evaluation efforts. Although we do not provide an exhaustive discussion of these issues here, the reader should be aware of the following limitations of this study:

- Because certain data (e.g., pre-post test scores) were not reported by all projects submitting evaluation reports, extrapolation of the findings based on those reporting useable data to those projects that did not report such data is not appropriate.

- The reader should note that the data reported and discussed here represents projects, not individual students or schools.

Despite the shortcomings of some evaluation reports, we found a great deal of evidence to suggest that the Colorado Bilingual programs have been effective in improving the English reading achievement of students in the program. This improved achievement in English reading should be of great interest to parents of children in these programs.